REPORT DOCUMENTATION PAGE

AD-A209 855

1b RESTRICTIVE MARKINGS	
2b DECLASSIFICATION / DOWNGRADING SCHEDULE	3 DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-IR- 89-0911
6a NAME OF PERFORMING ORGANIZATION: University of California	6b OFFICE SYMBOL (If applicable)
6c ADDRESS (City, State, and ZIP Code): La Jolla, CA 92093	7a NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research/NL
	7b ADDRESS (City, State, and ZIP Code): Building 410, Bolling AFB, DC 20332-6448
8a NAME OF FUNDING / SPONSORING ORGANIZATION: AFOSR	8b OFFICE SYMBOL (If applicable): NL
	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-86-0062
8c ADDRESS (City, State, and ZIP Code): Building 410, Bolling AFB, DC 20332-6448	10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO
61102F	2312	A1	

11 TITLE (Include Security Classification): The Back Propagation Technique For Modeling Cortical Computation

12 PERSONAL AUTHOR(S): Dr. David Zipser

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Final	FROM 01 Feb 88 TO 01 Jan 89	Jan 31, 1989	19

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Over the past several years powerful learning procedures have been developed that can program simulated neural networks to compute a wide variety of functions. This has made it possible to use learning procedures to train model networks to do computations that occur in the brain. While there was no a priori reason to suppose that the individual neuro-like units in these model networks would resemble the brain in any way, the empirical observation is that they do. Good results have been achieved applying this paradigm to modeling monkey parietal area 7a. Various aspects of the primary visual area have also been successfully modeled using this approach. The results of this work raise the interesting possibility that learning procedures and particularly the back propagation algorithm used in these studies, can serve as a general technique to account for how the brain implements computations. While these observations do not imply that back propagation is actually used in the brain, they do raise the possibility that some analogous learning procedure is used there.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	UNCLASSIFIED

22a NAME OF RESPONSIBLE INDIVIDUAL	22b TELEPHONE (Include Area Code)	22c OFFICE SYMBOL
Dr. William O Berry	(202) 767-5021	NL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

AFOSR-TR- 89-0911

FINAL TECHNICAL REPORT

AFOSR-86-0062

David Zipser

Institute for Cognitive Science

University of California, San Diego

La Jolla, California

09 MAY 1989

THE BACK PROPAGATION TECHNIQUE FOR MODELING CORTICAL COMPUTATION

I. Introduction

Over the past several years powerful learning procedures have been developed that can program simulated neural networks to compute a wide variety of functions. This has made it possible to use learning procedures to train model networks to do computations that occur in the brain. While there was no a priori reason to suppose that the individual neuron-like units in these model networks would resemble the brain in any way, the empirical observation is that they do. The response properties of some units in these networks closely resembles those of real neurons in the cortex. We have had particularly good results applying this paradigm to modeling monkey parietal area 7a (1,2,3,4). Various aspects of the primary visual area have also been successfully modeled by us and others using this approach (5,6). The results of this work raise the interesting possibility that learning procedures, and particularly the back propagation algorithm used in these studies, can serve as a general technique to account for how the brain implements computations. While these observations do not imply that back propagation is actually used in the brain, they do raise the possibility that some analogous learning procedure is used there.

II. Work Accomplished

A. Accounting for the Experimental Data From Parietal Area 7a

Lesions to the posterior parietal cortex in monkeys and humans produce profound spatial deficits in both motor behavior and perception (8,9,10,11). Based on single unit recording data and lesion studies Andersen and colleagues (12,13,14) proposed that parietal area 7a performed a spatial transformation from observation based to head centered coordinates by combining retinal based and eye-position information. Our model attempts to account for the mechanism of this transformation (1,2,3,4).

The classes of area 7a neurons that are relevant for our modeling effort are eye-position neurons, responding to eye-position only; visual neurons, responding to visual stimulation only; and spatially tuned neurons, which respond to both visual and eye position information. Neurons in the first two classes presumably represent the eye-position and retinal location information in observation based coordinates used by area 7a as input. The partially tuned neurons correspond to the hidden units in our model.

To model area 7a we used an input format based on experimental observations. The input consisted of two parts: an eye position and a retinal position. The output was a representation of spatial location. Training consists of randomly picking a set of allowed eye and retinal positions as input, and then computing the corresponding spatial location to train the output. This model, shown in Figure 1, is described in detail elsewhere (2). This network, as well as all others described here, was simulated using the P3 parallel system programming environment, implemented on a Sybolics 3600 LISP Machine (15).

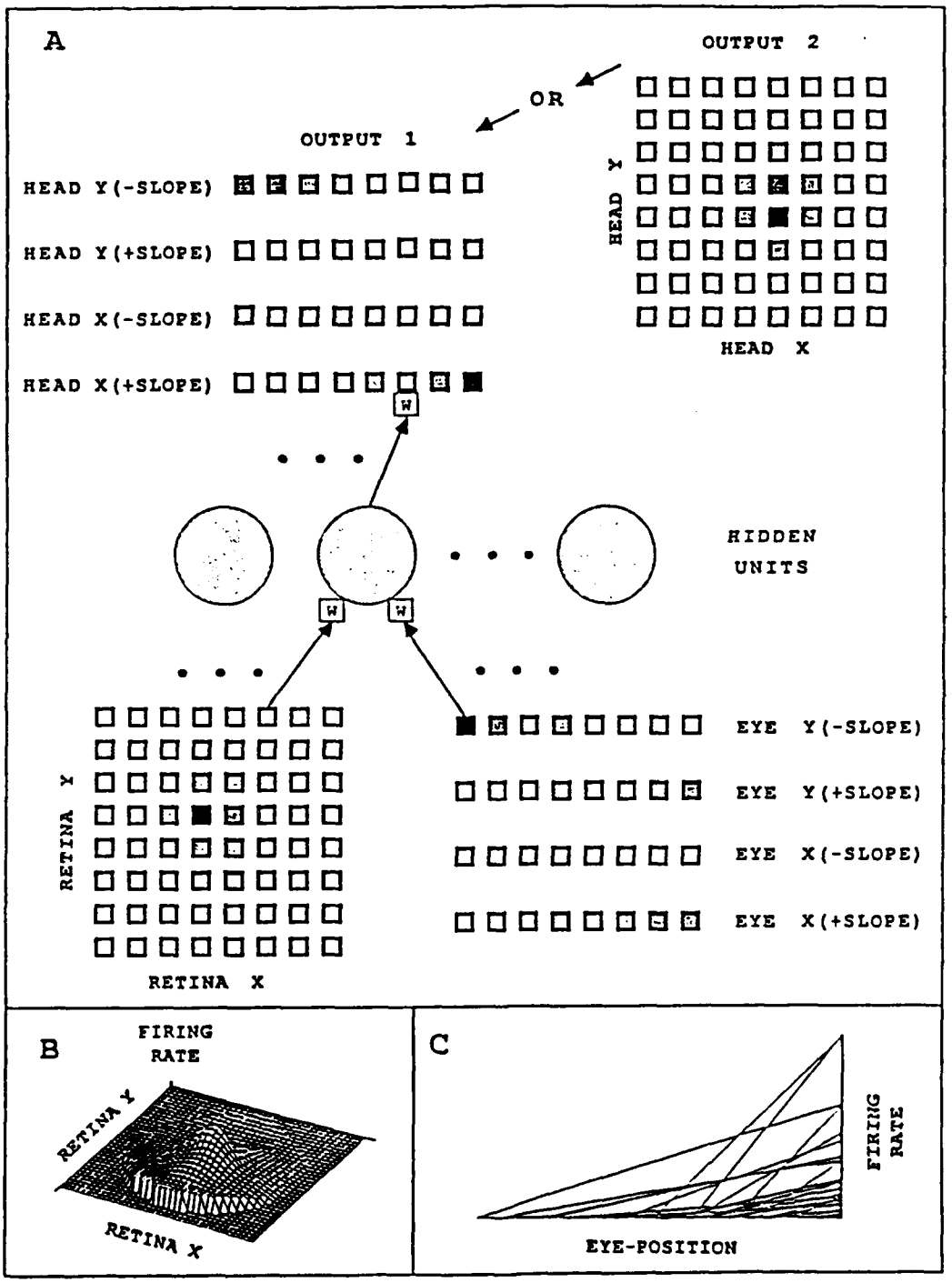

Figure 1. (a) The back propagation network used to simulate area 7a of the monkey parietal lobe. The input to the network consists of retinal position and eye-position information. The arrows indicate the direction of activity propagation; error was propagated back in the opposite direction. The *w*'s are the weights that are changed by learning. (b) An area 7a visual neuron receptive field of the type used to model input to the network in (a). (c) The eye position vs. firing rate response lines for 30 area 7a neurons showing the observed range of slopes and intercepts. The eye position input to the model was based on this data.

A teacher is needed for the network to learn the coordinate transformation carried out by area 7a. In our original studies we used two teacher formats, each representing spatial location in a head-centered frame. One format represented spatial location as the eye position at which the stimulus would be foviated. The other format represented head centered spatial location as the retinal location of the stimulus when looking straight ahead. These teachers were used to train the model in separate training sessions. Subsequently we have used a wider range of teachers and shown that any teacher format that encodes information about the head-centered location of the stimulus produces hidden units of the kind found in area 7a (4).

The network learned to compute the transformation carried out by area 7a. The really interesting observation is that the network learned to do this in a way analogous to area 7a. This can be seen by comparing the hidden units to the area 7a spatially tuned neurons. The important properties to be compared are retinal receptive fields and spatial gain fields. The extensive similarity of experimental and model receptive fields can be seen Figure 2. The spectrum of hidden unit receptive fields was similar for both teachers. Note that three of the most complex hidden unit receptive fields in row C of Figure 2(b) come from untrained hidden units. Training tends to smooth out the receptive fields and accentuate a single peak, usually shifting it toward the periphery.

Comparing the spatial gain fields is more complex because the interaction of observations taken in the presence and absence of visual stimulation must be considered. The eye position gain fields generated by the model are compared to the data for 7a neurons in Figure 3. All the total gain fields (outer circles) of the hidden units, generated by either teacher, were

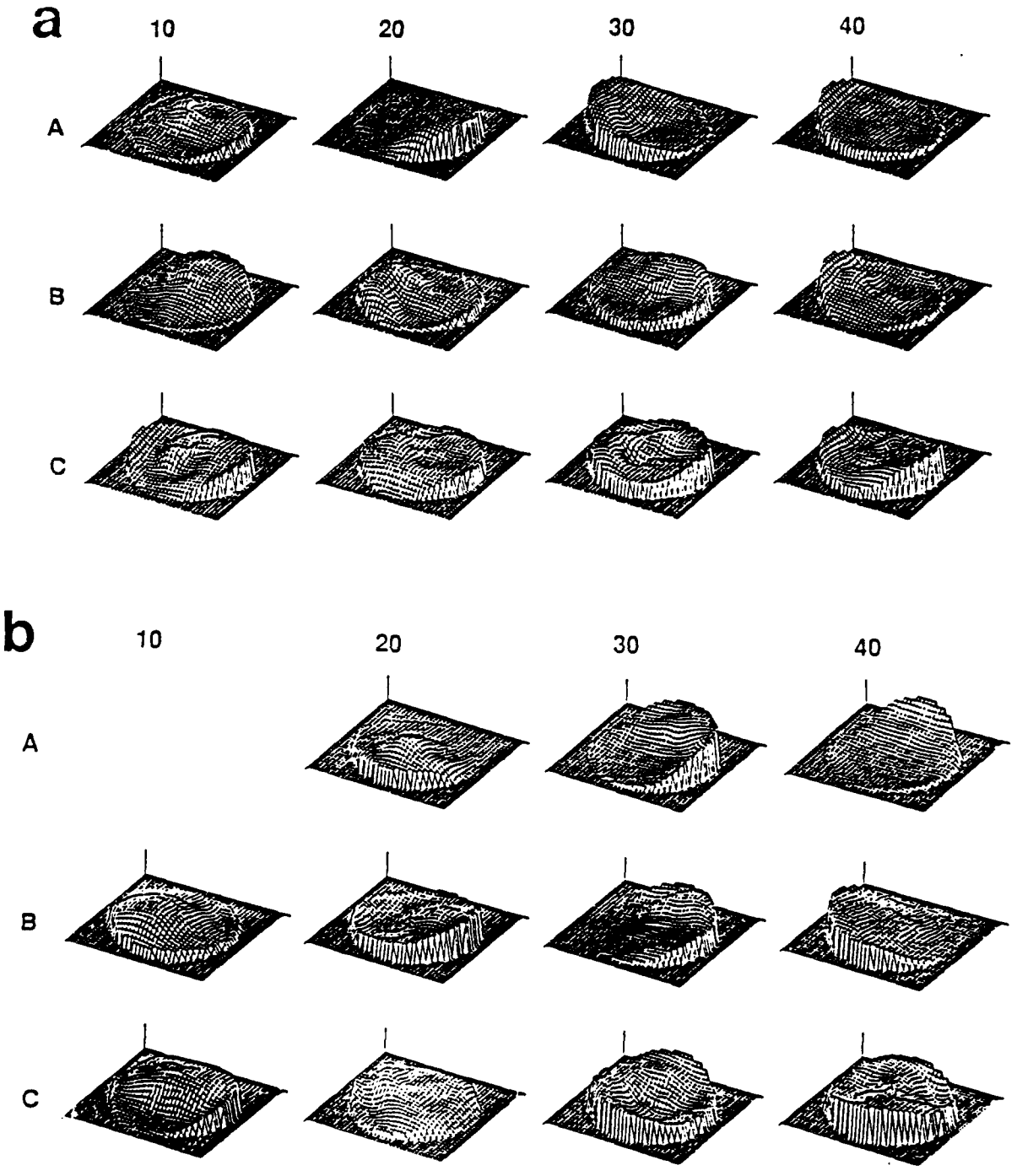

Figure 2. (a) Experimentally determined retinal receptive fields (1,6). The data for drawing each of these receptive field plots comes from measurements of the firing rate of a single area 7a neuron at 17 different retinal locations. These locations were at the center and at 10, 20, 30, and 40 degrees out. A neighborhood-weighted Gaussian smoothing function was used to create the plots shown here. The receptive fields are arranged in rows with the eccentricity of the field maxima increasing to the right, and in columns with the complexity of the fields increasing downward. All the fields in row A have single peaks. Those in B a few distinguishable peaks. The fields in C are the most complex. The data has been normalized so the highest peak in each field is the same height. (b) Hidden unit retinal receptive fields generated by the back propagation model. These plots were generated in the same way as those of Figure 2(a) except that the data came from computer simulations of the model network. All the fields, except for the three on the left in row C, are from units that have received 1,000 learning trials. The remaining three are from untrained units and represent fields that result from the initial random assignment of synaptic weights.

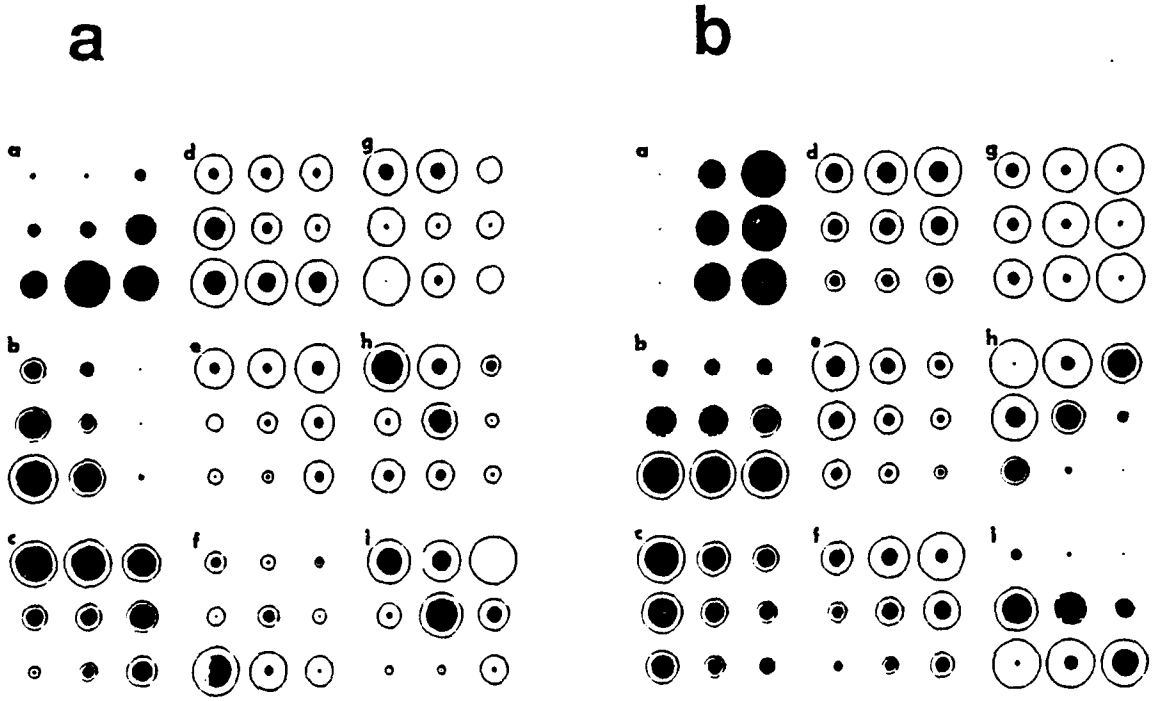

Figure 3. (a) The spatial gain fields of nine neurons from area 7a (1). The diameter of the darkened inner circle, representing the visually evoked gain fields, is calculated by subtracting the background activity recorded 500 msec before the stimulus onset from the total activity during the stimulus. The outer circle diameter, representing the total response gain fields, corresponds to the total activity during the stimulus. The annulus diameter corresponds to the background activity that is due to an eye position signal alone, recorded during the 500 msec prior to the stimulus presentation. (b) Hidden unit spatial gain fields generated by the model network. Fields *a-f* were generated using the monotonic format output; the rest used the Gaussian format output.

planar. This result compares with 80% for the 7a neurons. The visually evoked gain fields (inner dark disks) of the hidden units show differences between the teachers. With the monotonic eye-position format, 78% of the visual response gain fields were planar or monotonic, while with the Gaussian retinal format, only 36% fall in this class. These figures compare with 55% in this class for the experimental data. The striking similarity between model and experiment raises important questions such as to what degree will this result generalize to other cortical regions, and is there a back-propagation learning mechanism in the brain? Ultimately both of these are empirical questions that must be answered by more research.

B. Back Propagation Models of the Visual System

Back propagation requires some form of teacher. We have tried to use teachers that could be obtained locally from signals present in the cortical area being modeled. One way to do this is to have the teacher the same as the input. In this case the network learns to do an identity map which re-creates the input pattern on the output (16). What happens in identity mapping can be roughly described as a principle component analysis using the same number of dimensions as there are hidden units, followed by a rotation of coordinates so that the variance is distributed equally on each axis. The net effect of this, when the number of hidden units is less than the number of inputs, is that the hidden units have a lower dimensional encoding of the input than the original input pattern (17,18,19). Our experience gained using identity mapping in speech recognition and image compression (19,20) shows that this often leads to the representation of important stimulus features in a very efficient way. Identity mapping is a powerful correlation-based learning technique. It has been demonstrated that even weaker correlational procedures such as simple Hebbian learning can generate some of the properties of visual neurons (21). When we applied identity mapping to simplified models of the visual system we found hidden units that encoded stimulus location, depth and orientation (6). These results are summarized below.

1. Hidden unit identity map encoding of stereo depth. The visual cortex contains many binocular neurons, some of which are involved in extracting depth information from stereo images on the two retinas. To see how hidden units would encode depth information, a simplified binocular model of the visual system was used in an identity mapping study. To

avoid the complexities of three dimensions, linear retinas were used with depth as the second dimension. The method for determining disparity and the identity mapping network are shown in Figures 4 and 5. On each training cycle, a location was picked at random, within a circle around the fixation point. This location was then projected to the retinas through the focal point of each eye. The activity for each unit of the retinal array was computed as a Gaussian function of its distance from the location of this projection point. The depth of the chosen location, relative to the fixation point, is encoded in the disparity between its location on the two retinas. The flat, linear retinas used here are an approximation to a horizontal slice through the curved retinas of the eye.

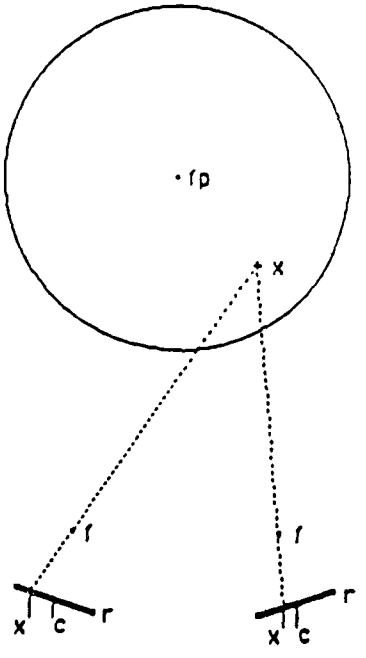

Figure 4. Method for computing disparities during training of two retina identity mapping network. The fixation point is indicated by *fp*, *c* is the center of the retinas, *x* the location of the stimulus in space and its projection on the retinas, and *f* is the focal point for each eye.

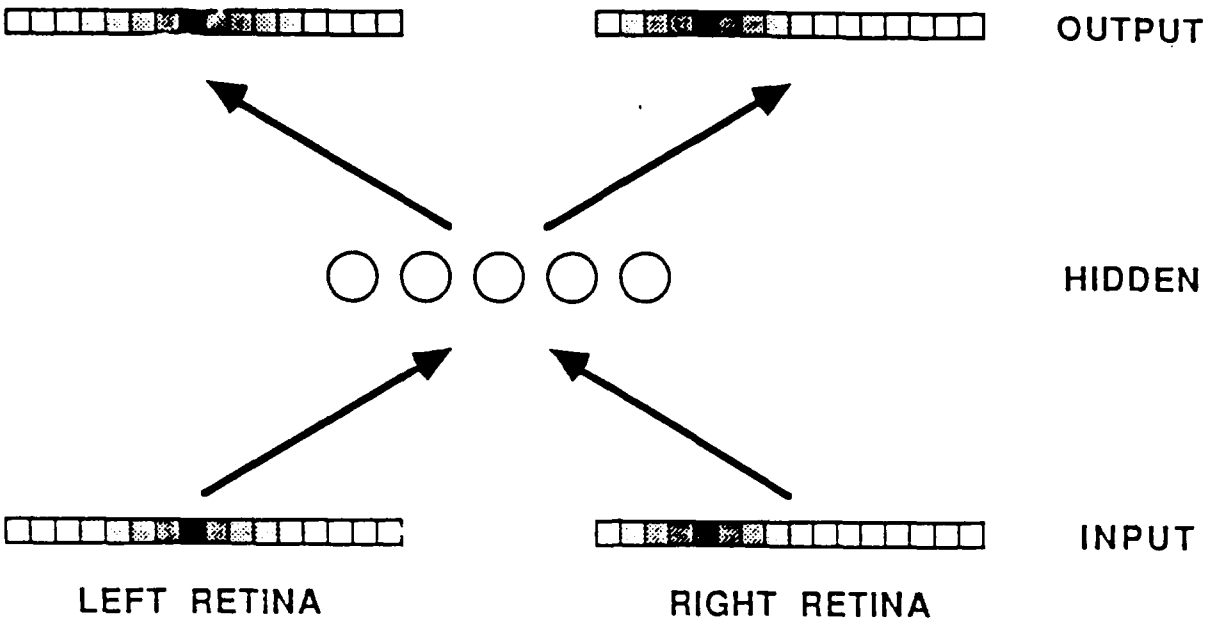

Figure 5. The two retina identity mapping network. All input units are connected to all hidden units, which are connected to all output units.

It is not completely trivial to determine what features the hidden units encode. A graphic display that shows the relationship between unit activity and spot position has proven very useful in this regard. To generate this type of display the input spot is scanned over the disk of possible input positions. At every position of the spot the activity of each hidden unit is plotted at a corresponding position on the computer display. A separate disk-shaped pattern is generated for each hidden unit. The pixels in the display used can be set only to black or whit, so to get a graded effect the pixel closest to the spot position is set to black with a probability approximately proportional to the unit's activity. To clearly define the disk the probability of setting a pixel to black is slighdy greater than zero even when activity is zero. This produces a display in which the degree of darkining over an area is roughly proportional to unit activity. To more clearly show the activity pattern, a set of contour lines is superimposed on the display by plotting white pixels whenever the activity of a unit falls within certain evenly

spaced narrow bands. The response of the hidden units as a function of disparity is discussed in the research plan section.

The results obtained from computer runs with two, three, four, and nine hidden units are shown in Figure 6. The network with only two hidden units could not solve the problem. In this network the input stimuli were mapped to identically corresponding positions on the two output arrays, that is, there was no disparity. For three and more hidden units the disparity present in the input was re-created to some degree on the output. The accuracy with which

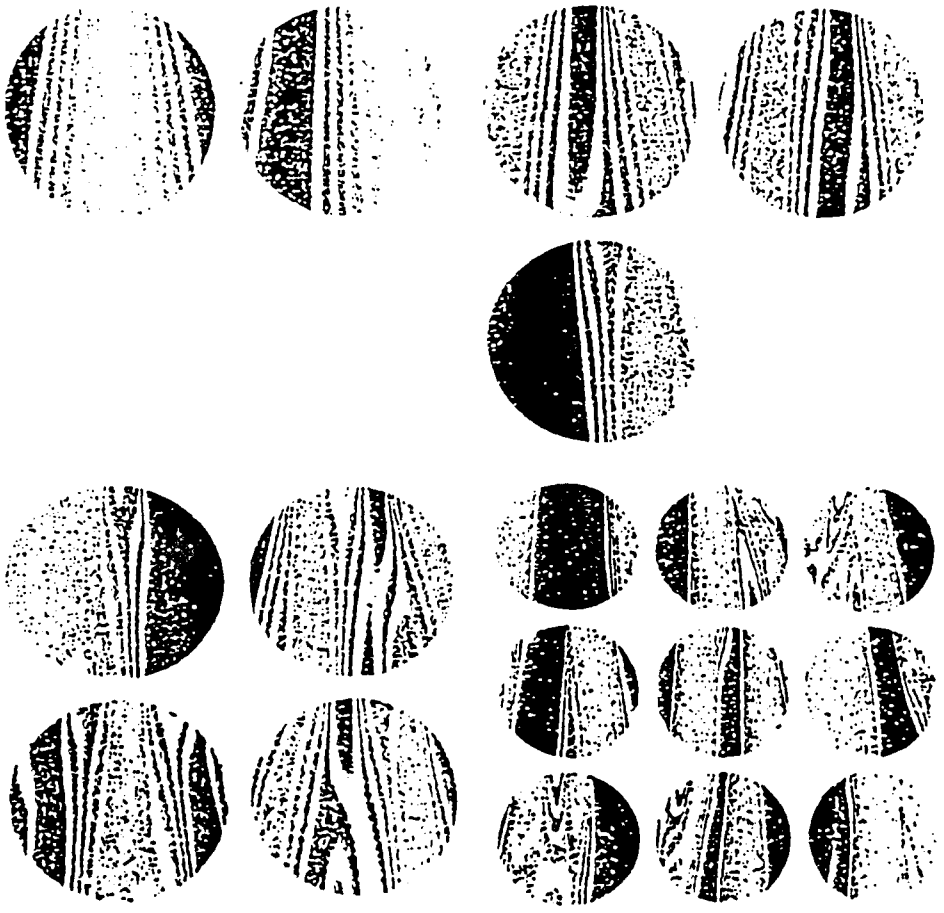

Figure 6. Hidden unit activities plotted as a function of stimulus location in space for the two retina identity mapping network. Runs using two, three, four, and nine hidden units are shown.

disparity was reproduced was quite good with four units and virtually perfect with nine units. Close examination of the data in Figure 6 reveals several interesting points. In the case of two hidden units the contour lines in the two eyes are completely parallel to each other. This means that no depth information can be derived from the activity of the hidden units because their activity values remain proportional for all depths at any given lateral displacement. The contour lines are angled in depth to compensate for the additional lateral movement required to produce a fixed displacement on the retina as a point gets further from the observer. For more than two hidden units the contour lines are no longer exactly parallel. This provides a coordinate system in depth that the output units can use in re-creating the required disparity. In the case of nine hidden units, sections of the contour lines are running nearly perpendicular to each other, providing detailed depth information.

In no case did the network solve the problem by treating the two eyes separately. For example, in the case of four hidden units, a simple solution would be to dedicate two of the hidden units to each eye. What is actually observed is a distributed, binocular representation for each hidden unit. The reason that monocular units did not occur is that back propagation identity mapping works by generating hidden units that capture the correlations in the input patterns. When, instead of the relatively small amount of disparity that results from depth, a large amount of random uncorrelated disparity is used in the training, we find that the hidden units do become monocular. This is analogous to the loss of binocular neurons in animals with defects preventing eye convergence.

2. *Hidden unit encodings of location and orientation.* When vision is used to guide spatial behavior, an important early step in the process is to extract the location of the stimulus from the retinal image. It is known from neurophysiological studies that retinal location serves as input to the parietal region (see Figure 1(b)), but the way the retinal location of a stimulus is computed is not understood. One possibility is that the hidden units in a network trained to identity map images of discrete objects will encode their location in the image.

To test this hypothesis a network embodying a very simple model of a visual system was used. The input and output layers of this network consisted of identically configured image arrays as required for identity mapping. Rather than using complex images, the initial studies were done with pattern sets consisting of circular Gaussian spots appearing at random locations on the input image plane. These spots are simplified examples of the images of discrete objects as they would appear at low resolution after figure ground separation.

A network typical of those used is shown in Figure 7. It consists of input and output layers each with 100 units arranged as 10×10 arrays. The hidden layer has a small number of units ranging from one to nine in different runs. The locations to be coded are represented as spots of activity on the input layer. The position of the center of a spot is picked at random on each learning cycle. To reduce edge effects the center of the spot is limited to a disk 8 units in diameter centered on the input array. The activity of each unit in the input array is determined as a Gaussian function of its distance from the center of the spot. Gaussian spots are used so that spot location could vary continuously, and also to allow a smooth, nondiscrete representation of the input stimuli. Training consists of applying a spot to the input and using this same spot as the target output. The job of the network is to learn to copy the input spot

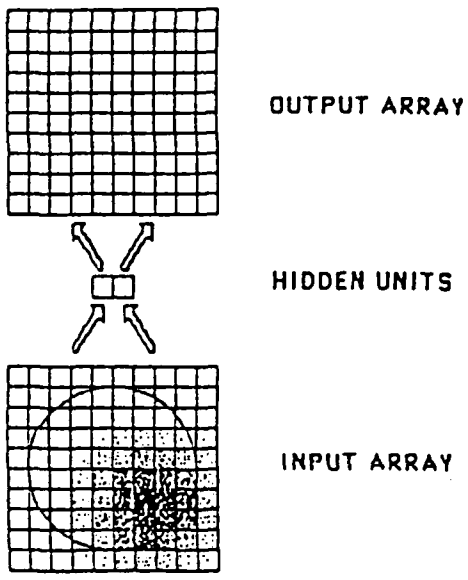

Figure 7. Identity mapping network for a single two-dimensional retina.

onto the output layer through the small number of hidden units.

Figure 8 shows results obtained after extensive identity map training using a network of the kind described above, with two hidden units. The two hidden units in Figure 8 can be seen to have formed an explicit encoding of spot location using an orthogonal coordinate system. Although the activity profiles of the two units are everywhere orthogonal they are not linear. This is presumably a consequence of constraints other than spot location such as spot size and spot shape. When the training is repeated with the same network using different sets of random starting weights and training sequences, the activity patterns learned by the hidden units are the same except for the angle of rotation of the coordinate axis. This angle is a free parameter and its final value is a result of unpredictable details of random parameters in the training procedure. When a different size Gaussian spot is used for training, the pattern of hidden unit activity is similar but there are some subtle differences in the curvature and separation of the

Figure 8. The response of two hidden units from an identity mapping network of the kind shown in Figure 7 plotted in retinal space. The center of the retina is at the center of the circles.

contour lines. This indicates that spot size information is also being coded by the hidden units.

The shape of the activity patterns that develop in the hidden units indicate that they might have an orientation tuning similar to neurons in the visual cortex. This was tested by using a progressively rotated dark bar as input to the fully trained networks. The result is shown in Figure 9, which is a plot of hidden unit activity as a function of the orientation of a Gaussian bar having the same width as the spots used to train the network. The pair of hidden units have a strong orientation tuning. It is interesting that one unit has a decrease in activity at the preferred orientation while the other has an increase. The preferred orientations of the two units are just 90 degrees out of phase. These observations on identity mapping of simple visual system models serve as the basis for our plans to develop a more complete model of visual cortex.

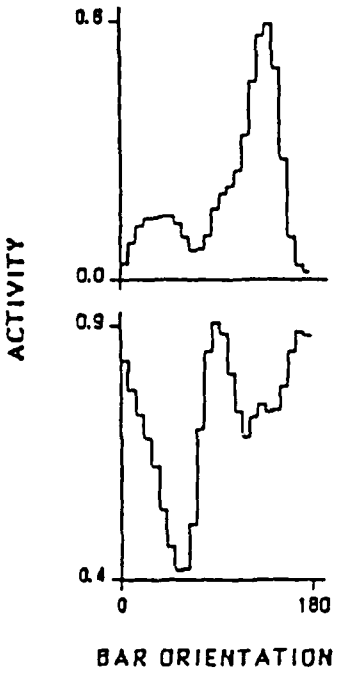

Figure 9. The response of a pair of hidden units to a rotating bar. The hidden units are those shown in Figure 8. After training was completed learning was turned off and the network was stimulated by an input that consisted of a gaussian bar passing through the center of the input array. The 1/*e* width of the bar was 0.2 of the array width, the same as the 1/*e* radius of the spot used for training. The bar was presented at 25 equally spaced orientations between 0 and 180 degrees. The figure shows the activity of the hidden units as a function of bar orientation.

References

1. Zipser, D., & Andersen, R. A. (1987). A network model using back propagation learning simulates the spatial tuning properties of posterior parietal neuron. *Society for Neuroscience Abstracts, 13*, pt. 2, 1452.

2. Zipser, D., & Andersen, R. A. (1988). A back propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature.*

3. Andersen, R. A., & Zipser, D. (in press). The role of the posterior parietal cortex in coordinate transformations for visual-motor integration. *Canadian Journal of Physiology and*

Pharmacology.

4. Zipser, D., & Andersen, R. A. (in press). The role of the teacher in learning-based models of parietal area 7a. *Brain Research Bulletin.*

5. Lehky, S. R., & Sejnowski, T. J. (1987). Extracting 3-D curvatures from images of surfaces using a neural model. *Society for Neuroscience Abstracts, 13*, pt. 2, 1451.

6. Zipser, D. (in press). Programming neural nets to do spatial computations. In N. E. Sharkey (Ed.), *Review of cognitive science I.* Norwood, NJ: Ablex.

7. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1. Foundations* (pp. 318-362). Cambridge, MA: MIT Press/Bradford Books.

8. Critchley, M. (1953). *The parietal lobes.* New York: Hafner.

9. Lynch, J. C. (1980). The functional organization of posterior parietal association cortex. *Behavioral Brain Sciences, 3*, 485-534.

10. Andersen, R. A. (in press). The neurobiological basis of spatial cognition: role of the parietal lobe. In J. Stiles-Davis, M. Kritchevsky, & U. Bellugi, (Eds.), *Spatial cognition: Brain bases and development.* Chicago: University of Chicago Press.

11. Bock, O., Eckmiller, R., & Andersen, R. A. (1987). *Goal-directed arm movements in trained monkeys following ibotenic acid lesions in the posterior parietal cortex.* Manuscript submitted for publication.

12. Andersen, R. A., Siegel, R. M., & Essick, G. K. (in press). Neurons of area 7 activated by both visual stimuli and oculomotor behavior. *Experimental Brain Research.*

13. Andersen, R. A., & Mountcastle, V. B. (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *Journal of Neurosciences, 3,* 532-548.

14. Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science, 230,* 546-548.

15. Zipser, D., & Rabin, D. (1986). P3: A parallel network simulating system. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1. Foundations..* Cambridge, MA: MIT Press/Bradford Press.

16. Ackley, D., Hinton, G. E., & Sejnowski, T. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science, 9,* 147-169.

17. Hinton, G. E. Personal communication.

18. Saund, E. (1987). *Dimensionality-reduction using connectionist networks* (AI Memo 941). Cambridge: Massachusetts Institute of Technology, AI Laboratory.

19. Cottrell, G. W., Munro, P. W., & Zipser, D. (in press). Image compression by back propagation: A demonstration of extensional programming. In N. E. Sharkey (Ed.), *Review of cognitive science I.* Norwood, NJ: Ablex. (Also ICS Tech. Rep. No. 8702. La Jolla: University of California, San Diego, Institute for Cognitive Science.)

20. Elman, J. & Zipser, D. (in press). Learning the hidden structure of speech. *Journal of the Acoustical Society of America.* (Also ICS Tech. Rep. No. 8701. La Jolla: University of California, San Diego, Institute for Cognitive Science.)

21. Linsker, R. (1987). From basic network principles to neural architecture: Emergence of orientation selective cells. *PNAS, 83,* 8390-8394.